IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 13, 2019, accepted August 28, 2019, date of publication September 11, 2019, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2940211

Virtualization-Based Efficient TSV Repair for 3-D Integrated Circuits

MUHAMMAD IMRAN[1], (Student Member, IEEE), HYUNSEUNG HAN[2], JOOHO KIM[1], (Student Member, IEEE), TAEHYUN KWON[3,4], (Student Member, IEEE), JAEYONG CHUNG[5], (Member, IEEE), AND JOON-SUNG YANG[6], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[2]Memory Division, Samsung Electronics, Hwaseong 18448, South Korea
[3]Department of Semiconductor and Display Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[4]System LSI Division, Samsung Electronics, Hwaseong 18448, South Korea
[5]Department of Electronics Engineering, Incheon National University, Incheon 22012, South Korea
[6]Department of Systems Semiconductor Engineering, Yonsei University, Seoul 03722, South Korea

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT), through the AI Graduate School Support Program, under Grant 2019-0-00421, in part by the Basic Science Research Program through the National Research Foundation of Korea by the Ministry of Education under Grant NRF-2018R1D1A1B07049842, in part by the Ministry of Trade, Industry Energy (MOTIE) under Grant 10080594, in part by the Korea Semiconductor Research Consortium (KSRC) Support Program for the Development of the Future Semiconductor Device, in part by the Competency Development Program for Industry Specialists of the Korean Ministry of Trade, Industry and Energy (MOTIE), Operated by the Korea Institute for Advancement of Technology (KIAT) under Grant N0001883, and in part by the HRD Program for Intelligent semiconductor Industry.

ABSTRACT Three-dimensional (3-D) integration offers a promising solution to the technology scaling barriers. Reliability of the 3-D Integrated Circuits (ICs) is highly dependent on the integrity of the underlying interconnect. Through Silicon Via (TSV) based 3-D ICs would suffer from low yield due to the faults in TSVs. In addition, TSVs introduce stress and noise in the substrate. Adding redundant TSVs for repairing the faulty ones has commonly been proposed to improve the yield of TSV-based 3-D ICs. The existing TSV repair approaches employ a number of spare TSVs in a group of signal TSVs. We propose a TSV virtualization based repair architecture which utilizes a single redundant TSV to repair multiple faulty TSVs. The proposed architecture relies on transmitting multiple bits through a single TSV using multi-level voltage quantization. It makes efficient use of the TSV redundancies in repairing the faulty TSVs. With less number of spare TSVs, the proposed architecture can reduce the area overhead by more than 70%. Reduction in the TSV count allows greater interconnect density and helps to mitigate the TSV-induced noise and stresses. Alternatively, for a similar number of spare TSVs, the proposed method can enhance the fault tolerance capability of the conventional approaches thus leading to an enhanced chip yield. The eye diagram simulations using an electrical model of the TSV show a reduction of less than 5% in noise margin when using a 16-level voltage quantization at a data rate of 5 Gbps which is typical for 3-D integration applications.

INDEX TERMS 3-D integrated circuit, eye diagram simulations, multiple bits transmission, through silicon via (TSV), TSV Virtualization, TSV repair.

I. INTRODUCTION

Semiconductor device scaling is aimed at performance enhancement by reducing the interconnect delays, increasing density and decreasing power dissipation [1]. Technology scaling is limited by several factors such as process variations, tunneling current leakage, increasing parasitics and interconnect power dissipation constraints [2], [3]. These limitations are hampering the continuation of the technology scaling trend. Three dimensional (3-D) integration offers a viable solution to overcome some of these problems and continue the integration trend [4]–[6]. 3-D integrated circuits (ICs) come with a variety of benefits such as higher density, reduced interconnect delays and low power

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian Zambelli.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

operation. In addition, the 3-D ICs can help meet the increasing input-output (I/O) pins demand. Memory systems are becoming increasingly critical to the computing performance especially with the introduction of multi-core processor architectures. 3-D ICs, with short interconnects and more channels, can help achieve greater memory bandwidth to keep pace with the other factors in improving computing performance [5], [6].

The yield of 3-D ICs depends on the integrity of all the required TSVs. As presented in [7], with the increase in TSV count, there is a significant reduction in yield. The TSV fabrication process introduces tensile stresses in silicon and can result in undesired stress fields [8]. In addition to utilizing the silicon area, TSVs can be a major source of noise in the substrate which increases with the reduction in spacing between the TSVs [9]. The fabrication of TSVs requires additional components such as landing pads and bumps. All these factors limit the overall number and density of the TSVs.

TSVs are vulnerable to several faults introduced during the manufacturing process. These include misalignment, added impurities and random open defects [10], [11]. The pre-bonding test prior to die stacking is used to prevent the yield loss due to stacking of the faulty dies [12], [13]. A post-bonding test is used to verify the stacked 3-D IC. Although the probability of failure in an individual TSV is low [14], however, without a repair mechanism, a single defective TSV can lead to the failure of an entire chip. Therefore, several TSV repair mechanisms have been proposed to improve the yield of 3-D ICs, thus saving the cost incurred due to yield loss [14]–[19].

The existing TSV repair architectures use multiple redundant TSVs to replace the faulty ones. In these architectures, a group of signal TSVs along with some redundant TSVs are used. The proportion of redundant TSVs in a group depends on the reliability of the manufacturing process. Not all the redundant TSVs are utilized in the repair process and many of the TSVs are reduced to mere additional metal lines in the substrate, a fact known as the inefficiency in TSV redundancy utilization [20]. Keeping in view the challenges posed by TSV based 3-D integration technology, we propose a TSV virtualization based repair architecture utilizing multi-bit transmission through a single TSV channel. With the proposed method, a single TSV can be used to repair multiple defective TSVs in a group. The proposed architecture offers a low cost and efficient TSV repair solution while reducing the overall TSV count, thus allowing greater interconnect density. Eye diagram simulations are performed to evaluate the impact of the proposed multi-level voltage quantization on the noise margins. Compared to the existing TSV repair architectures, the proposed architecture is efficient in utilizing the spare TSVs and also requires less area overhead.

The remainder of the paper is organized as follows. Section II summarizes existing TSV repair techniques. The proposed TSV repair architecture is presented in Section III. In Section IV, we present an evaluation of the proposed architecture with results from the eye diagram simulation along with a comparison of the proposed method with conventional TSV repair architectures. Section V concludes the paper.

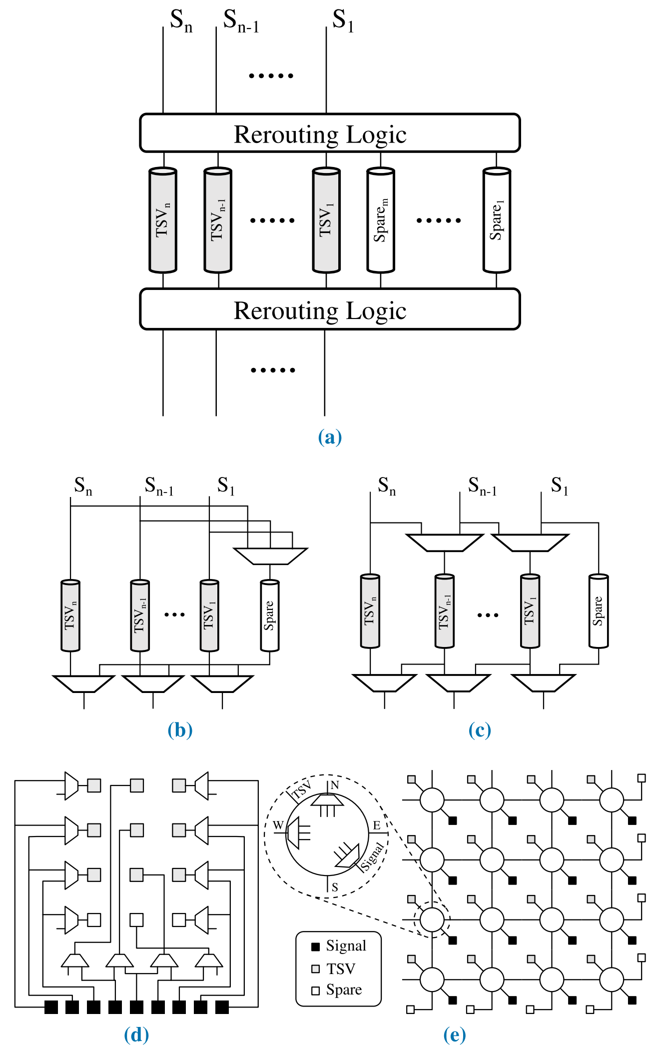

FIGURE 1. Existing TSV repair techniques. (a) Generic repair architecture. (b) Signal switching [15]. (c) Signal shifting [14]. (d) Crossbar [7]. and (e) NoC type repair architecture [16].

II. PREVIOUS WORKS

For TSV repair, most of the existing works require hardware redundancy architectures to enhance the yield of 3-D ICs [14]–[19], [21]. In these approaches, spare TSVs are included along with the signal TSVs and the faulty TSVs can be replaced with the spare ones to improve the overall yield during testing. The ratio of the spare TSVs to the signal TSVs in a group is decided based on the acceptable yield and the quality of the manufacturing process. A more reliable manufacturing process requires less number of spare TSVs (for a given number of signal TSVs) to be specified in design [20]. Figure 1(a) shows a generic TSV repair architecture consisting of n signal TSVs with m spares. The difference in various repair schemes lies in the rerouting architectures.

Rerouting can be implemented using either the fuse technology or switching [20].

TSV repair architectures can broadly be classified into either one dimensional (1-D) or two dimensional (2-D). In the 1-D repair schemes, TSVs are arranged in the form of a chain. Two popular approaches using this scheme are the signal switching [15] and signal shifting [14], shown respectively in Figure 1(b) and Figure 1(c). Switching is based on rerouting of the faulty TSVs to the spare ones without involvement of other working TSVs. In shifting, all the TSVs in group are shifted from fault location to the spare TSV. Signal switching requires less overhead for the rerouting logic and involves rerouting only the faulty TSVs but leads to greater timing imbalance.

In the 2-D repair approaches, the TSVs are arranged in the form of a grid. In [7], a 2-D crossbar architecture has been proposed, in which the signal TSVs and the spare TSVs are arranged in the form of a crossbar. Either rows or columns of spare TSVs are added in this approach. As the Figure 1(d) shows, a row of spare TSVs has been added. Through the crossbar, every spare TSV is linked to the signal TSVs in the corresponding column and it can repair any faulty TSV in that column by shifting the signals from fault location in the direction of the spare TSV. Another popular repair technique, among 2-D approaches, is a network-on-a-chip (NoC) type architecture proposed in [16], in which the faulty TSVs are repaired using the redundant TSVs that are farther apart. Signal rerouting to a spare TSV is achieved through a network of switches which are implemented using multiplexers. The NoC type repair architecture is shown in Figure 1(e). As shown in the figure, a switch can route the signal to either the working TSV in the absence of a fault or towards either East (E) or South (S) direction where redundant TSVs are located. This kind of repair approach is particularly useful for repairing the clustered faults since the faulty TSVs can be replaced by a group of spares that are located at a distance. However, the hardware overhead in this approach is large and many signals get re-routed even without faults leading to additional delays. In general, the 2-D repair techniques lead to greater yield, however, it may require more area overhead and routing complexity.

All the existing TSV repair approaches use multiple spare TSVs. The fact that many of the spare TSVs are not utilized highlights the inefficiency in existing approaches. Based on this observation, the authors in [20] proposed an online testing and soft error detection architecture for further enhancing the reliability of 3-D integrated circuits by utilizing the redundant TSVs that are left unused in the repair process. In this architecture, response compactors at the input and output of the TSVs are used for error detection. The redundant TSVs are used for transmitting the input side compactor's response to the output side for comparison.

Efficient utilization of the TSV redundancies and reduction in the TSV count are desirable to mitigate the TSV induced stresses and noise while still enhancing the overall yield. Less number of repair TSVs also allow for more signal TSVs to be incorporated in a 3-D IC, thus increasing the interconnect density. With this motivation, we propose the TSV virtualization based repair approach which can provide the same amount of repair capability with the reduced number of spare TSVs. With proposed architecture, the redundant TSVs are efficiently utilized. Additionally, the soft error detection mechanism proposed in [20] can still be applied for further enhancing the reliability of 3-D ICs.

A recent work [22] has introduced an adaptive fault-tolerance structure generation method. In this method, a TSV repair structure is adaptively generated. It considers varying the fault-tolerance of TSV repair groups and assigning spares from a large number of candidates while minimizing the hardware-cost and efficiently utilizing the spare TSVs. The proposed vritualiztion-based TSV repair concept is applicable to the adaptive fault-tolerance structure generation methods such as [22] for more efficient utilization of the spare TSVs. As such, the proposed method enhances the fault-tolerance capability of an individual spare TSV and is, therefore, orthogonal to the conventional techniques which focus on effective spare TSVs placement topologies.

III. THE PROPOSED ARCHITECTURE

TSV virtualization is an idea of utilizing a single TSV to serve the purpose of multiple interconnects between different dies. It involves multi-bit transmission of data using a single physical channel. This technique can be used for low overhead TSV repair. In Section III-A, the idea of multi-level voltage quantization for TSV virtualization is discussed. The proposed repair architecture has been described in Section III-B. In order to evaluate the impact of proposed quantization technique with respect to noise margins, this work performs eye diagram simulations using an electrical model of the TSV. Section III-D describes the TSV model used for signal analysis.

A. MULTI-LEVEL VOLTAGE QUANTIZATION BASED TSV VIRTUALIZATION

TSV Virtualization for multi-bit transmission can be realized by using a digital to analog converter (DAC). The DAC converts n bit data into a single analog signal having n levels of voltage quantization. In this way, n-bit information is transmitted using a single TSV channel. On the receiving side, an analog to digital converter (ADC) converts the analog signal back to n bit data. Figure 2(a) shows the architecture for TSV virtualization using DAC and ADC and Figure 2(b) illustrates the idea of voltage quantization to send two-bit information using 4 levels. With the proposed TSV virtualization concept, a single TSV can be used in place of multiple TSVs. This helps to reduce the TSV count and improve the utilization of spare TSVs in a redundancy-based TSV repair architecture.

B. TSV VIRTUALIZATION BASED REPAIR ARCHITECTURE

The proposed TSV virtualization scheme helps in reducing the overall number of TSVs compared to the conventional

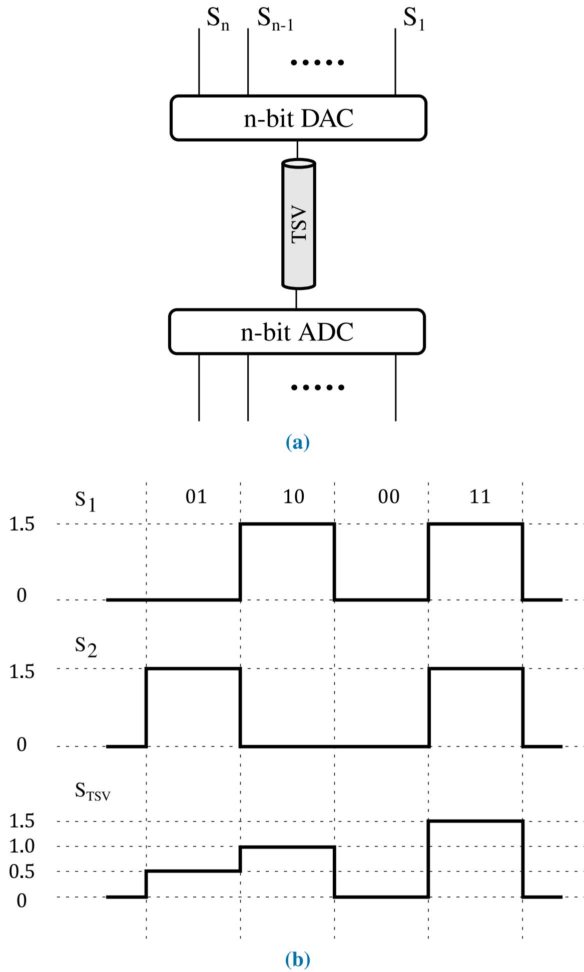

FIGURE 2. **(a) TSV Virtualization for *n* bit transmission. (b) Voltage quantization for 2 bit data (S_{TSV} is the transmitted through TSV instead of S_1 and S_2).**

approach. In the proposed repair architecture, the TSV count can be adjusted based on the application at cost of varying overhead. Figure 3(a) illustrates the generic architecture for virtualization based TSV repair. The comparison between the proposed architecture (Figure 3(a)) and the conventional one (Figure 1(a)), reveals the potential benefit of the proposed method. The requirement of m spare TSVs in previous approaches translates into just one TSV and m bit resolution of the DAC and ADC used for TSV virtualization. This leads to utilizing one spare TSV for repairing up to m failing TSVs in group of n TSVs. The rerouting logic for repairing the faulty TSVs can be implemented by using the pass transistor logic, fuse technology or multiplexers-based signal switching [16], [20], [23]–[25]. After identifying the defective TSVs, they can be rerouted to the spare TSV, thus improving the overall yield of the 3-D ICs. Unlike the conventional approaches, in the proposed architecture, there's less probability of a redundant TSV being left unused. However, with the presence of multi-bit DACs and ADCs, any online testing mechanism which makes use of the residual repair resources, such as that proposed in [20], can still be applied for enhanced reliability.

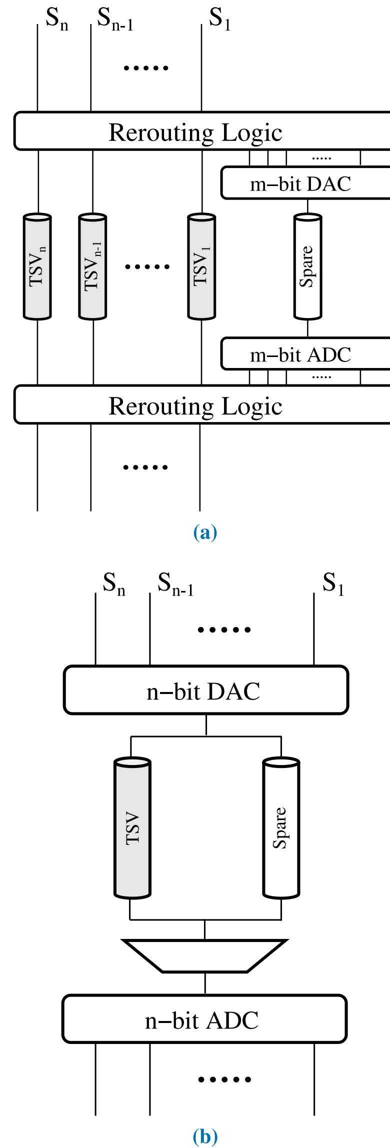

FIGURE 3. **TSV Virtualization based repair. (a) Overall architecture. (b) Specific case where number of signals = resolution of DAC / ADC and 100% TSV redundancy.**

The proposed architecture can be utilized in any of the configurations of the prior TSV repair architecture to address various kinds of faults including both the clustered faults and distributed random faults. The corresponding overhead for the rerouting logic would also be similar. The main benefit of the proposed architecture is a reduced number of spare TSVs and improved spare TSVs utilization in a repair architecture. However, the use of multi-level voltage quantization may impact the noise margins in the transmitted signals. Therefore, it is important to analyze the impact of proposed TSV virtualization on signal quality.

C. CHOOSING AN OPTIMAL CONFIGURATION

The choice of an appropriate value of n (TSVs in a group) and m (resolution of DAC and ADC) depends on the probability of defective TSVs and the tolerable overhead for implementing

the DAC and ADC. This might vary depending on the application. Figure 3(b) shows an ideal scenario in which n equals m and the proposed method is used to reduce the number of signal TSVs as well as the spares. The shown example case uses only two TSVs for n signal TSVs and has a higher degree of reliability due to 100% TSV redundancy. It also uses a simple routing logic with only one 2-1 multiplexer. However, in some applications, it may not be economical to use a single TSV to replace a large group of signal TSVs at cost of high-resolution DAC and ADC circuit. In general, higher value of m increases the area overhead for implementation of corresponding DAC and ADC.

Another important factor that limits the value of m is the reduction in noise margin for the transmitted signal due to the multi-level voltage quantization. The noise margin also depends on the transmitted data rate and the TSV channel characteristics.

Eye diagram simulations are frequently used in communications to analyze the channel response in the presence of noise and inter-symbol interference. An eye diagram is obtained by overlapping various bit transitions. Such a pattern can be used for measuring noise margins and various timing parameters. In this paper, to analyze the decrease in noise margin by the proposed quantization technique, eye diagram simulations are performed using an electrical model of the TSV. A test input consisting of random bit sequence with random transitions to different quantization level is used. Eye diagrams are obtained by overlapping plots of the corresponding output response. The decrease in noise margins is measured by using the minimum eye openings corresponding to different responses. The electrical model of the TSV used for these simulations is discussed next.

D. TSV MODELING

To analyze the response of the TSV channel to an input signal, electrical modeling of TSV has been actively researched. Modeling is also important for designing of I/O channels in modern 3-D integrated circuits. Various equivalent electrical models of the Through Silicon Via have been proposed in the literature [26]–[29]. To evaluate the impact of the proposed method on noise margins, we use the electrical model of TSV proposed in [26] which is experimentally verified. It's a high frequency scalable electrical model which considers most of the parasitic components associated with the TSV as well as the interconnecting bumps and redistribution layer (RDL). The bumps act as joints for the TSV while the purpose of redistribution layer is to distribute the signals over different locations [26]. As these are essential components in 3-D ICs, it is necessary to consider them in electrical modeling of the TSV. By using the analytic *RLGC* equations for the given model, the value of different patristics can be obtained. Additionally, using the equivalent electrical circuit, the TSV channel response to a given input can be analyzed.

To verify the response of TSV channel with the proposed quantization technique, we consider a ground and signal TSV pair along with a redistribution layer (RDL). Figure 4(a) and 4(b) respectively show the cross-sectional and the side view of the TSV and RDL schematic used in simulation. Figure4(c) illustrates the detailed dimensions of the TSV, bump and RDL. These dimensions have been adopted from [26]. The equivalent electrical circuit for this model is

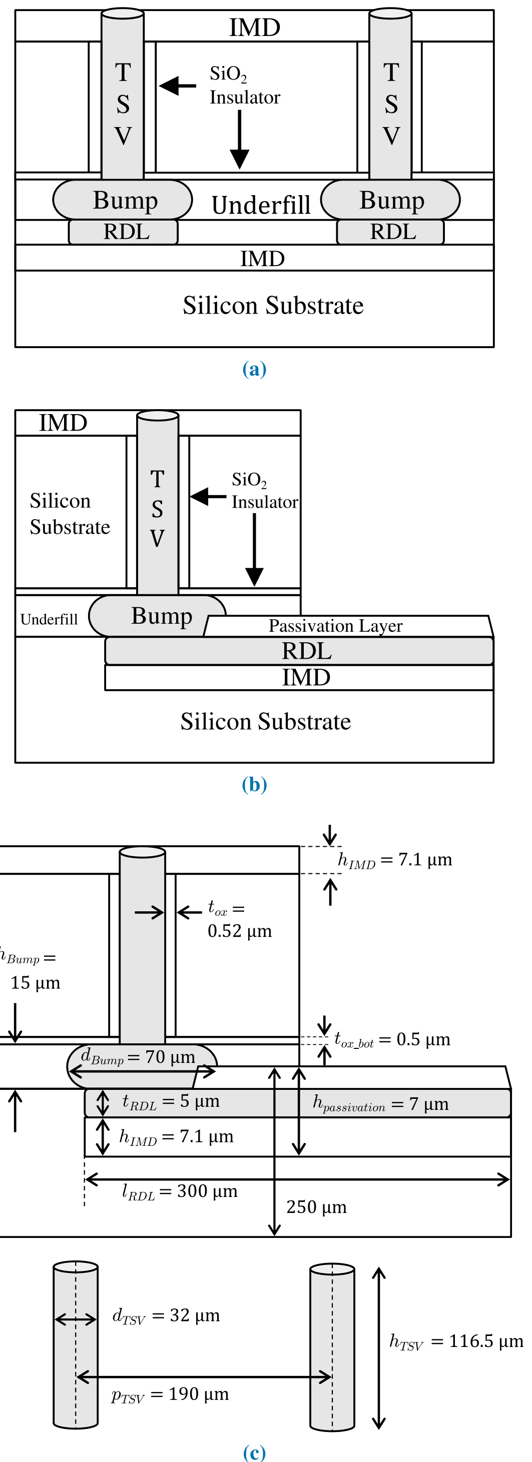

FIGURE 4. TSV and RDL test model. (a) Cross-sectional view, (b) side view and (c) detailed dimensions for TSVs, Bumps and RDL interconnect in the test model (based on [26]).

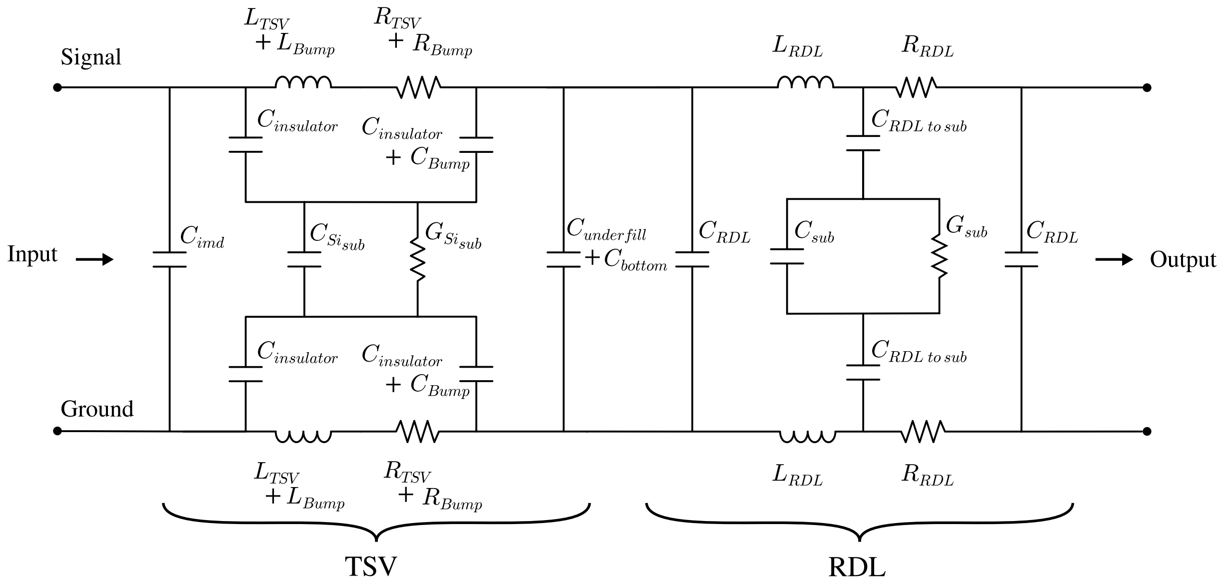

FIGURE 5. Equivalent circuit for test model of TSV and RDL.

shown in Figure 5. The modeling of various parameters of the TSV using equivalent electrical components, is summarized next.

TSVs are electrically isolated from the silicon substrate through a silicon dioxide (SiO_2) insulation layer. This is equivalently modeled by $C_{insulator}$ capacitance which can be expressed using the coaxial-cable capacitance model. At the bottom end of the TSV, the capacitance between the bump and substrate, represented as C_{Bump}, is added to the $C_{insulator}$. C_{Bump} can be derived using parallel plate capacitance model. The two capacitances namely the C_{imd} and the C_{bottom}, represent the capacitances between the signal and ground TSVs through the top inter-metal dielectric (IMD) layer and the bottom silicon dioxide (SiO_2) insulation layer. These are modeled using the parallel wire capacitance model. Using the same principle the capacitance between the bumps through the underfill layer is modeled as $C_{underfill}$. The capacitance and conductance between the TSVs through the silicon substrate are modeled as C_{Sisub} and G_{Sisub}, respectively, using the parallel wires model. The resistances of both the TSV and the bump are given as R_{TSV} and R_{Bump}, respectively. The derivation of these resistances takes the skin effect due to alternating current into consideration. At high frequencies, the inductance impedance becomes prominent, therefore it is necessary to consider the inductance of the TSV and bump modeled as L_{TSV} and L_{Bump}.

In a manner similar to that described for the TSV, various parameters associated with the redistribution layer (RDL) are modeled as C_{RDL}, L_{RDL}, R_{RDL}, C_{sub} and the G_{sub}. Material properties used in the calculations have also been adopted from [26].

The aforementioned electrical model is used for the eye diagram simulations (discussed in next section) to evaluate the impact of proposed voltage quantization on noise margins.

IV. EVALUATION

We use eye diagram simulations to confirm the viability of the proposed architecture and to analyze the reduction in signal noise margin by the proposed multi-level voltage quantization. We also evaluate the various aspects such as yield, number of TSVs and area overhead for the proposed method compared to the conventional TSV repair approaches.

A. EYE DIAGRAM SIMULATIONS

Eye diagrams are helpful in analyzing a communication channel's response. The signal degradation and noise margin can easily be estimated from the eye patterns. For the proposed method of multi-level voltage quantization, eye diagram simulations have been performed to measure the expected reduction in noise margin. To obtain the eye diagrams, the electrical model shown in Figure 5 is used. With the dimensions given in Figure 4(c), the values of various circuit elements are calculated by using the *RLGC* equations in [26]. The test input consists of random bit sequence of 20 Kbits. Four different transmission schemes are considered i.e. 1-bit, 2-bit, 3-bit and 4-bit using 2, 4, 8 and 16 levels of quantization, respectively. The input eye diagrams have been shown in Figure 6. These eye diagrams are obtained by overlapping multiple input transitions. For each input to the test model of the TSV, there's a corresponding output transition. Output eye patterns are obtained by overlapping the output responses. Figure 7 shows the corresponding output eye diagrams against the test input. Different diagrams are obtained by varying the input data rate between 1 Gbps, 5 Gbps and 10 Gbps. By measuring the minimum eye opening in each case, the corresponding noise margin is estimated.

As Figure 7 shows, the noise margin from eye diagrams decreases as multiple bits are transmitted by using more number of levels for voltage quantization. Furthermore, it is

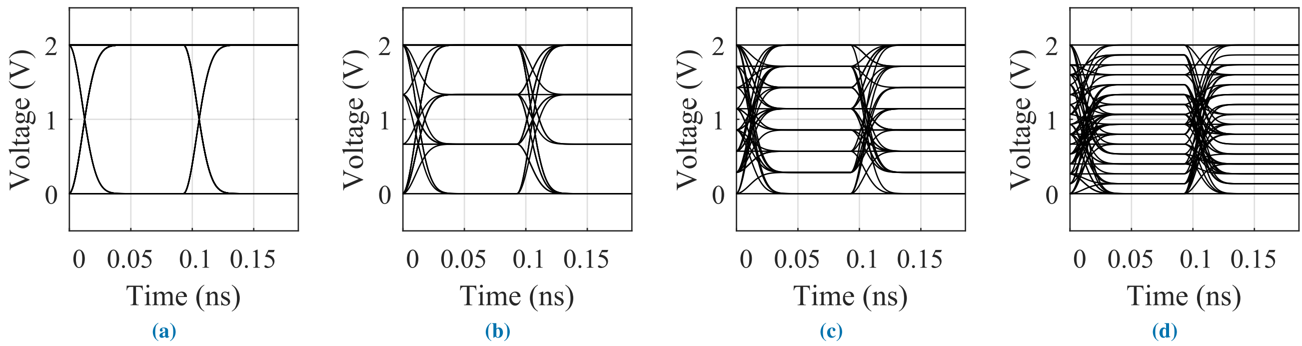

FIGURE 6. Eye diagrams showing the input test sequence (@ 10 Gbps) using (a) two level (b) four level (c) eight level and (d) sixteen level voltage quantization.

FIGURE 7. Eye diagrams showing the output response (@ 1 Gbps, 5 Gbps and 10 Gbps) using (a) two level (b) four level (c) eight level and (d) sixteen level voltage quantization.

observed that the reduction in noise margin is proportional to the input data rate. The measured values of minimum eye openings for different cases are listed in Table 1. The reduction in noise margin when employing multi-bit transmission, compared to the single bit transmission using 2 levels, is also listed in the table. By varying the input data rate as well as the V_{dd} voltage level, different measurements are obtained. As seen from Table 1, for the input data rate of 1 Gbps, the reduction in worst-case noise margin remains below 0.25% as the V_{dd} voltage varies from 1.2 V to 2.1 V. The results also show that altering the V_{dd} voltage does not significantly impact the noise margins. A slight improvement in noise margin is observed for higher values of V_{dd}. Increasing the data rate, however, causes a more significant decrease in the noise margin. At the input data rate of 10 Gbps, the eye opening and hence the noise margin decrease from about 97.5% for 2-level voltage quantization to a minimum of 61.3% for 16-level voltage quantization. Therefore,

TABLE 1. Noise margin measurements from output eye diagrams.

Vdd Voltage	Data Rate	Two level quantization	Four level quatization		Eight level quatization		Sixteen level quatization	
		Eye-opening	Minimum Eye-opening	Margin Reduction	Minimum Eye-opening	Margin Reduction	Minimum Eye-opening	Margin Reduction
1.2 V	1 Gbps	1.1997 V (99.97%)	0.3998 V (99.95%)	0.025%	0.1712 V (99.88%)	0.094%	0.0798 V (99.74%)	0.232%
	5 Gbps	1.1961 V (99.67%)	0.3964 V (99.09%)	0.581%	0.1678 V (97.90%)	1.773%	0.0764 V (95.48%)	4.193%
	10 Gbps	1.1702 V (97.52%)	0.3693 V (92.32%)	5.198%	0.1415 V (82.57%)	14.950%	0.0491 V (61.33%)	36.189%
1.5 V	1 Gbps	1.4996 V (99.97%)	0.4997 V (99.95%)	0.025%	0.2140 V (99.88%)	0.094%	0.0997 V (99.74%)	0.232%
	5 Gbps	1.4951 V (99.67%)	0.4955 V (99.09%)	0.581%	0.2097 V (97.86%)	1.812%	0.0955 V (95.54%)	4.135%
	10 Gbps	1.4628 V (97.52%)	0.4616 V (92.32%)	5.198%	0.1762 V (82.22%)	15.296%	0.0628 V (62.78%)	34.739%
2.1 V	1 Gbps	2.0994 V (99.97%)	0.6996 V (99.95%)	0.025%	0.2996 V (99.88%)	0.094%	0.1396 V (99.74%)	0.232%
	5 Gbps	2.0931 V (99.67%)	0.6937 V (99.10%)	0.577%	0.2936 V (97.87%)	2.104%	0.1337 V (95.51%)	4.16%
	10 Gbps	2.0479 V (97.52%)	0.6462 V (92.32%)	5.198%	0.2473 V (82.43%)	15.084%	0.0904 V (65.59%)	32.924%

the maximum reduction in noise margin is about 36.2% at a data rate of 10 Gbps. Considering 4-bit transmission with 16 voltage levels, the noise margin reduction remains below 5% for the data rate up to 5 Gbps, which shows possibility of transmitting even more bits at this data rate.

For a practical application, we consider the TSV-based 3-D DDR4 SDRAM presented in [30]. The IO speed of the 3-D SDRAM is 2.4 Gbps at a supply voltage of 1.2 V. Based on our experimental results, for this memory, the proposed TSV repair architecture can easily be incorporated with a 16-level voltage quantization with less than 5% reduction in noise margin. With a data rate below 5 Gbps, even higher levels of voltage quantization can be used to reduce the overall redundant TSVs and enhance the efficiency of the repair architecture.

The results from eye diagram simulations confirm that the proposed multi-level voltage quantization technique is viable for efficient TSV repair. The reduction in noise margins is within acceptable range for practical implementations. The number of levels used, depends significantly on the data rate. The other factors include characteristics of the TSV channel and the acceptable overhead for DAC and ADC. Therefore, based on application, different variants of the proposed TSV repair architecture may be considered.

B. COMPARISON WITH CONVENTIONAL TSV REPAIR ARCHITECTURES

In comparison with the conventional schemes, the proposed virtualization-based TSV repair architecture decreases the number of spare TSVs for a given level of reliability. Alternatively, for the same number of spare TSVs, by increasing the virtualization, the proposed architecture can have increased fault tolerance capability. However, if the TSV count is kept same, the proposed architecture would require additional cost of ADC and DAC. Different topologies for placing and routing spare TSVs, such as the ones shown in Figure 1, can be considered with the proposed virtualization method. This architecture can cover both the clustered faults as well as the random faults. Considering a similar overhead for rerouting logic, we next compare the proposed architecture with conventional approaches in terms of yield and hardware cost. Since the proposed architecture can be used in any of the configurations of the existing repair architectures, we compare the proposed method with the conventional methods in terms of number of TSVs, chip yield and area of TSVs and DACs / ADCs.

1) YIELD ANALYSIS WITH RESPECT TO NUMBER OF TSVs

The proposed architecture can reduce the TSV count for a similar level of TSV repair capability as the conventional TSV repair architectures. Alternatively, with a fixed number of signal and spare TSVs, the proposed method can improve the fault tolerance of the repair architecture by using multi-level voltage quantization. Increasing the number of faulty TSVs handled by a repair approach can enhance the chip yield. To analyze the impact of proposed multi-level voltage quantization on chip yield, we consider a probabilistic model. The failure rate of a TSV has been reported to be 10^{-4} to 10^{-5} [20]. Considering a failure rate of 10^{-4}, we analyze the chip yield for different schemes assuming that a single spare TSV is used for repair. The results are shown in Figure 8. In the figure, *Virtualization-n* indicates an n-bit virtualization scheme in which 2^n voltage quantization levels are used. Figure 8 shows that for a given number of TSVs, the proposed method can considerably improve the yield compared to conventional methods. For example, with Virtualization-4 scheme, the proposed method has more than 99% yield for over 12K TSVs. For the same number of TSVs, the conventional repair methods achieve only 63% yield (36% less). This analysis also shows that for a given level of yield requirement, the proposed method can allow more signal TSVs to be used by simply increasing the level

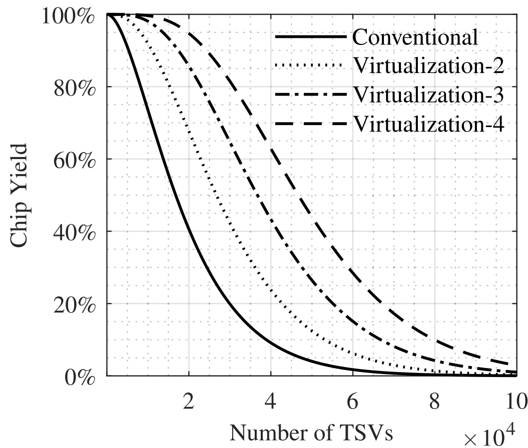

FIGURE 8. Comparison of chip yield and number of TSVs.

of voltage quantization. The area overhead by increasing the virtualization level will not be as significant as adding more spare TSVs with additional routing logic. For a given target yield, a *Virtualization-n* scheme requires *n* times less spare TSVs. The reduction in TSVs also reduces the area overhead as discussed next.

2) COMPARISON OF AREA OVERHEAD

Reduction in TSV count by the proposed virtualization architecture comes at cost of overhead for DAC and ADC circuits. We compare the chip area of conventional TSV repair architectures to that of the proposed architecture based on the prior works on practical TSV-based 3-D ICs and efficient implementations of ADC and DAC circuits. Exploring efficient implementations of a DAC or an ADC is beyond the scope of this research and there has been much research about efficient implementations of DAC and ADC [31]–[36]. We therefore refer to the existing works on practical implementations of DACs and ADCs. Although different technology nodes are considered for practical implementations, however, in comparing the area overhead, we adopt a conservative approach, in which 180nm technology is considered for the DAC circuit in the proposed architecture and a 50nm technology is considered for TSVs in the conventional architectures.

In practical 3-D ICs, TSVs are usually placed to form a grid with a minimum *pitch* (distance between two TSVs) requirement. Area overhead by the TSVs, therefore, depends on the pitch which varies depending on application. Considering a practical 3-D DDR3 DRAM using 50nm node DRAM process, presented in [15], the TSV pitch is 80μm. With this pitch, the footprint area overhead per TSV is 80 $\times$ 80μm^2. Considering the fact that the TSVs are normally arranged in a grid, the overall volume of a 3-D TSV grid can be even higher. Nevertheless, reducing 1 TSV leads to at least 80 $\times$ 80μm^2 area saving in each of the dies stacked around TSV, providing room for proposed DAC and ADC circuits. The silicon substrate around TSVs, which is wasted because of minimum pitch requirement, can otherwise be used for useful logic in stacked dies. Although a practical implementation of the proposed architecture may use multiple smaller resolution DACs and ADCs, yet for the sake of area comparison we consider a 10-bit DAC presented in [32]. The 10-bit DAC in [32], using 180nm process technology, has a total die area of 110 $\times$ 94μm. A 6-bit ADC based on [31], using 40nm process technology, has a die area of 0.00058mm^2. With a 10-bit DAC in the proposed architecture, 10 spare TSVs can be replaced by a single TSV. This would require an area overhead of 0.01674mm^2. The corresponding area overhead for 10 TSVs is 0.064mm^2, showing a reduction of 74% in area overhead by the proposed architecture. It should also be noted that 180nm technology for the DAC circuit is considered. The area overhead for DAC would be even less when considering the DRAM technology node which is below 65nm. A similar analysis shows that the 6-bit ADC circuit of [31] along with a single TSV requires 82% less area compared to 6 TSVs in a conventional repair architecture. The DAC and ADC area is compared separately because reducing each TSV leads to less footprint area in two stacked dies, which can be used for DAC and ADC circuit. Considering the area comparison based on 10-bit DAC, Figure 9(a) compares the area overhead of the proposed architecture against that of the conventional TSV repair architectures with increasing number of spare TSVs. The practical 3-D DDR3 DRAM [15] considered for area comparison has 150 spare TSVs. With 150 spare TSVs, the proposed architecture requires 74% less area overhead.

Area overhead for the conventional TSV repair architectures depends on the TSV pitch. The above comparison considers TSV pitch of 80μm, based on the practical DDR3 DRAM. However, if the pitch is very small, the area overhead for the proposed ADC and DAC may be dominant. Figure 9(b) compares the area overhead by varying the TSV pitch. In this comparison, the number of spare TSVs is fixed at 150 while the TSV pitch is varied from 10μm to 100μm. The results show that the proposed architecture reduces the chip area compared to conventional TSV repair architectures when the TSV pitch is more than 33μm. With more recent technology, TSV pitch could be even smaller to offset the hardware cost of ADC and DAC in the proposed architecture.

3) TIMING OVERHEAD

The use of DACs and ADCs in the proposed TSV repair architecture introduces additional timing overhead for conversion from digital to analog and analog to digital. However, for typical applications such as a 3-D memory, with the sampling rate of DACs and ADC of a few Gs/s, the latency would not be critical. This is especially true for the emerging memory technologies that have considerably higher access latencies compared to conventional memories. Moreover, the additional latency can be completely hidden when data buffering is used which is typical in practical memory systems.

4) POWER AND ENERGY OVERHEAD

Using the electrical equivalent model of TSV considered in this work, the power dissipation and propagation delay for

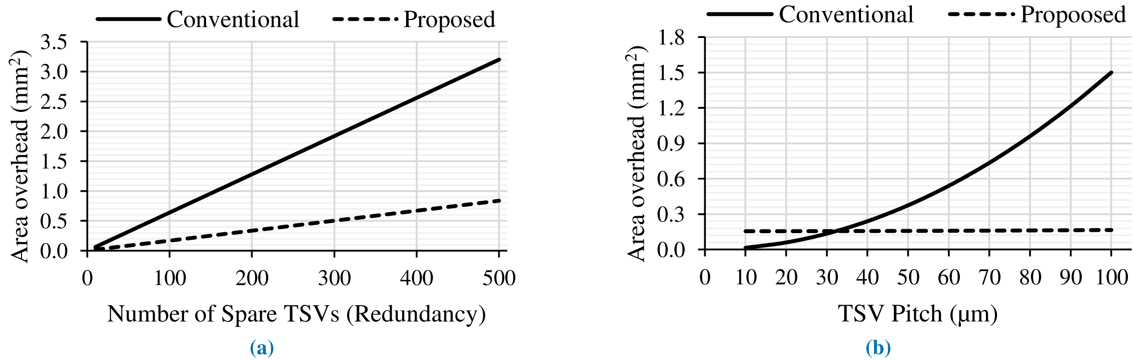

FIGURE 9. Comparison of area overhead (for similar yield) by varying (a) Number of Spare TSVs (b) TSV Pitch.

TSVs have been estimated in [37]. Based on these estimations, Cu-based TSVs dissipate a few μW power and the equivalent energy is of the order of $10^{-16} \sim 10^{-17}$J. Depending on implementation, ADC and DAC [31]–[36] have conversion energies ranging from a few fJ/conversion-step to a few tens of fJ/conversion-step. Although, the conversion energy of ADC and DAC is higher compared to energy dissipated by the TSVs, however, a part of this energy overhead is compensated by the reduction in number of TSVs (from hundreds to tens) using the proposed TSV repair architecture. Moreover, the energy overhead would be added only when the data is transferred between different dies of a 3-D IC and for those TSVs groups where faulty TSVs have been repaired.

Considering practical 3-D ICs such as high-density memories, depending on the memory technology, the per bit programming energy varies from a few fJ to tens of pJ [38]. Emerging memory technologies such as a multi-level cell Phase Change Memory can have per cell programming energy in hundreds of pJ [39]. Considering the overall programming energy for a large number of bits, the overhead of the ADC and DAC conversion energies would be negligible.

V. CONCLUSION

Hardware redundancy is the widely used approach to improve the yield of 3-D integrated circuits. Spare TSVs are used to replace the failing TSVs. The existing repair architectures require several redundant TSVs, thus causing greater TSV-induced stresses and noise. Many of the spare TSVs are not utilized during the repair process. We introduce a new TSV repair architecture based on the TSV virtualization to transmit multiple bits through a single TSV using multi-level voltage quantization. Using the proposed repair technique, a single redundant TSV can be used to repair multiple failing TSVs in a group. This leads to an efficient utilization of the TSV redundancy, thus avoiding the existence of unused spare TSVs in a working chip. It can help improve the interconnect density and also mitigate the TSV-induced stresses and noise. Depending on the application, data rate and the characteristics of the TSV channel, the resolution of voltage quantization can be adjusted. The results from eye diagram simulations show less than 5% decrease in noise margin at a data rate of 5 Gbps using 16-level voltage quantization. The proposed TSV repair architecture requires about 70% less area overhead compared to the conventional repair architectures.

REFERENCES

[1] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, Jul. 1999.
[2] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, Mar. 2001.
[3] D. J. Frank, "Power-constrained CMOS scaling limits," *IBM J. Res. Dev.*, vol. 46, nos. 2–3, pp. 235–244, 2002.
[4] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," *Proc. IEEE*, vol. 94, no. 6, pp. 1214–1224, Jun. 2006.
[5] K. N. Chen and C. S. Tan, "Integration schemes and enabling technologies for three-dimensional integrated circuits," *IET Comput. Digit. Techn.*, vol. 5, no. 3, pp. 160–168, May 2011.
[6] K. Tran, "Scaling the next-generation architecture: The integration of 2.5D and 3D integrated circuits," *IEEE Solid State Circuits Mag.*, vol. 8, no. 2, pp. 35–42, Jun. 2016.
[7] I. Loi, S. Mitra, T. H. Lee, S. Fujita, and L. Benini, "A low-overhead fault tolerance scheme for TSV-based 3D network on chip links," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, San Jose, CA, USA, Nov. 2008, pp. 598–602.
[8] P. Dilion and J. Mulder, "Through silicon via technology status," in *Proc. 3rd NASA Electron. Parts Packag. (NEPP) Electron. Technol. Workshop ETW*, Jun. 2012. [Online]. Available: https://nepp.nasa.gov/workshops/etw2012/talks/Tuesday/T08_Dillon_Through_Silicon_Via.pdf
[9] N. H. Khan, S. M. Alam, and S. Hassoun, "Through-Silicon Via (TSV)-induced noise characterization and noise mitigation using coaxial TSVs," in *Proc. IEEE Int. Conf. 3D Syst. Integr.*, Sep. 2009, pp. 1–7.
[10] M. Pathak, J. Pak, D. Z. Pan, and S. K. Lim, "Electromigration modeling and full-chip reliability analysis for BEOL interconnect in TSV-based 3D ICs," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2011, pp. 555–562.
[11] L. Jiang, Y. Liu, L. Duan, Y. Xie, and Q. Xu, "Modeling TSV open defects in 3D-stacked DRAM," in *Proc. IEEE Int. Test Conf. (ITC)*, Nov. 2010, pp. 1–9.
[12] B. Noia, S. Panth, K. Chakrabarty, and S. K. Lim, "Scan test of die logic in 3-D ICs using TSV probing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 2, pp. 317–330, Feb. 2015.
[13] C.-C. Chi, E. J. Marinissen, S. K. Goel, and C.-W. Wu, "Low-cost post-bond testing of 3-D ICs containing a passive silicon interposer base," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 22, no. 11, pp. 2388–2401, Nov. 2014.
[14] A.-C. Hsieh and T. Hwang, "TSV redundancy: Architecture and design issues in 3-D IC," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 4, pp. 711–722, Apr. 2012.
[15] U. Kang *et al.*, "8 Gb 3-D DDR3 DRAM using through-silicon-via technology," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 111–119, Jan. 2010.

[16] L. Jiang, Q. Xu, and B. Eklow, "On effective TSV repair for 3D-stacked ICs," in *Proc. DATE*, Mar. 2012, pp. 793–798.

[17] Y.-J. Huang and J.-F. Li, "Built-in self-repair scheme for the TSVs in 3-D ICs," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 31, no. 10, pp. 1600–1613, Oct. 2012.

[18] Q. Xu, S. Chen, X. Xu, and B. Yu, "Clustered fault tolerance TSV planning for 3-D integrated circuits," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 36, no. 8, pp. 1287–1300, Aug. 2017.

[19] W.-H. Lo, K. Chi, and T. Hwang, "Architecture of ring-based redundant TSV for clustered faults," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 12, pp. 3437–3449, Dec. 2016.

[20] J. Park, M. Cheong, and S. Kang, "R^2-TSV: A repairable and reliable TSV set structure reutilizing redundancies," *IEEE Trans. Rel.*, vol. 66, no. 2, pp. 458–466, Jun. 2017.

[21] J.-S. Yang, T. H. Han, D. Kobla, and E. L. Ju, "Dynamic self-repair architectures for defective through-silicon vias," *ETRI J.*, vol. 36, no. 2, pp. 301–308, 2014.

[22] S. Chen, Q. Xu, and B. Yu, "Adaptive 3D-IC TSV fault tolerance structure generation," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 38, no. 5, pp. 949–960, May 2019.

[23] C. Scholl and B. Becker, "On the generation of multiplexer circuits for pass transistor logic," in *Proc. Design, Automat. Test Eur. Conf. Exhib.*, New York, NY, USA, Mar. 2000, pp. 372–379. doi: 10.1109/DATE.2000.840298'

[24] J. MacArthur and T. M. Lacey, "Techniques and circuits for high yield improvements in programmable devices using redundant routing resources," U.S. Patent 5 925 920 A, Jul. 20, 1999.

[25] R. Zimmermann and W. Fichtner, "Low-power logic styles: CMOS versus pass-transistor logic," *IEEE J. Solid-State Circuits*, vol. 32, no. 7, pp. 1079–1090, Jul. 1997.

[26] J. Kim, J. S. Pak, J. Cho, E. Song, J. Cho, H. Kim, T. Song, J. Lee, H. Lee, "High-frequency scalable electrical model and analysis of a through silicon via (TSV)," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 1, no. 2, pp. 181–195, Feb. 2011.

[27] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, "Electrical modeling and characterization of through silicon via for three-dimensional ICs," *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 256–262, Jan. 2010.

[28] L. Cadix, A. Farcy, C. Bermond, C. Fuchs, P. Leduc, M. Rousseau, M. Assous, A. Valentian, J. Roullard, E. Eid, N. Sillon, B. Flechet, and P. Ancey, "Modelling of Through Silicon Via RF performance and impact on signal transmission in 3D integrated circuits," in *Proc. IEEE Int. Conf. 3D Syst. Integr.*, Sep. 2009, pp. 1–7.

[29] C. Ryu, J. Lee, H. Lee, K. Lee, T. Oh, and J. Kim, "High frequency electrical model of through wafer via for 3-D stacked chip packaging," in *Proc. 1st Electron. Systemintegr. Technol. Conf.*, vol. 1, Sep. 2006, pp. 215–220.

[30] R. Oh, B. Lee, S.-W. Shin, W. Bae, H. Choi, I. Song, Y.-S. Lee, J.-H. Choi, C.-W. Kim, S.-J. Jang, and J. S. Choi, "Design technologies for a 1.2V 2.4Gb/s/pin high capacity DDR4 SDRAM with TSVs," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2014, pp. 1–2.

[31] K. D. Choo, J. Bell, and M. P. Flynn, "27.3 Area-efficient 1GS/s 6b SAR ADC with charge-injection-cell-based DAC," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan./Feb. 2016, pp. 460–461.

[32] B. Greenley, R. Veith, D.-Y. Chang, and U.-K. Moon, "A low-Voltage 10-bit CMOS DAC in 0.01-mm/sup 2/ die area," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 52, no. 5, pp. 246–250, May 2005.

[33] K.-T. Lin, Y.-W. Cheng, and K.-T. Tang, "A 0.5 V 1.28-MS/s 4.68-fJ/Conversion-Step SAR ADC with energy-efficient DAC and trilevel switching scheme," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 4, pp. 1441–1449, Apr. 2016.

[34] Y.-H. Chung and C.-W. Yen, "An 11-bit 100-MS/s subranged-SAR ADC in 65-nm CMOS," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 12, pp. 3434–3443, Dec. 2017.

[35] H. Lee, A. Aurangozeb, S. Park, J. Kim, and C. Kim, "A 6-bit 2.5-GS/s time-interleaved analog-to-digital converter using resistor-array sharing digital-to-analog converter," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 11, pp. 2371–2383, Nov. 2015.

[36] H. Fan, D. Li, K. Zhang, Y. Cen, Q. Feng, F. Qiao, and H. Heidari, "A 4-channel 12-bit high-voltage radiation-hardened digital-to-analog converter for low orbit satellite applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 11, pp. 3698–3706, Nov. 2018.

[37] B. K. Kaushik, V. R. Kumar, M. K. Majumder, and A. Alam, *Through Silicon Vias: Materials, Models, Design, and Performance*. Boca Raton, FL, USA: CRC Press, 2016.

[38] S. Yu and P.-Y. Chen, "Emerging memory technologies: Recent trends and prospects," *IEEE Solid State Circuits Mag.*, vol. 8, no. 2, pp. 43–56, Jun. 2016.

[39] J. Wang, X. Dong, G. Sun, D. Niu, and Y. Xie, "Energy-efficient multi-level cell phase-change memory system with data encoding," in *Proc. IEEE 29th Int. Conf. Comput. Design (ICCD)*, Oct. 2011, pp. 175–182.

MUHAMMAD IMRAN received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2012. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Sungkyunkwan University, Suwon, South Korea. He was a recipient of the scholarship by the Higher Education Commission of Pakistan for his M.S. and Ph.D. studies. His research interests include very large-scale integration design and reliable architectures for the emerging memory technologies.

HYUNSEUNG HAN received the B.S. and M.S. degrees in semiconductor engineering from Sungkyunkwan University, South Korea, in 2014 and 2016, respectively. He has been with the Memory Division, Samsung Electronics, since 2014, as an Engineer, working on flash memory controller development. His research interests include memory systems, SoC design, and embedded systems.

JOOHO KIM received the B.S. degree in electronics engineering from Konkuk University, Seoul, South Korea, in 2017. He is currently pursuing the master's degree with the Department of Electrical and Computer engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include neural networks, memory systems, computer architecture, SoC design, and embedded systems.

TAEHYUN KWON received the B.S. degree in electrical engineering from Kyunghee University, South Korea, in 2005, and the M.S. degree in electrical engineering from Yonsei University, South Korea, in 2007. He is currently pursuing the Ph.D. degree in semiconductor and display engineering with Sungkyunkwan University, South Korea. He has been with the SoC Design Team, Samsung Electronics, South Korea. He is also a part of the team developing high performance CPU and GPU for Samsung mobile SoC products. His research interests include high-performance processor architecture, low power CPU and GPU implementations, high-bandwidth memory, and reliable architectures for emerging memory technologies.

JAEYONG CHUNG received the B.S. degree in electrical engineering from Yonsei University, Seoul, South Korea, in 2006, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Department of Electrical and Computer Engineering, University of Texas, Austin, TX, USA, in 2008 and 2011, respectively. He is currently an Assistant Professor with the Department of Electronic Engineering, Incheon National University, Incheon, South Korea. He was with the Strategic CAD Lab (SCL), Intel, and the IBM T. J. Watson Research Center, from summers of 2008 and 2010, respectively. From 2011 to 2013, he was with the Design Compiler Team, Synopsys Inc., Mountain View, CA, USA. His current research interests include neuromorphic systems and deep learning. He was a recipient of the Best Paper Award nominations at the International Conference on Computer-Aided Design (ICCAD), in 2009, and the Asia and South Pacific Design Automation Conference (ASPDAC), in 2011. One of his coauthored articles was selected in the Asian Test Symposium (ATS) 20th Anniversary Compendium.

JOON-SUNG YANG received the B.S. degree in electrical and computer engineering from Yonsei University, Seoul, South Korea, in 2003, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2007 and 2009, respectively. He was with Intel Corporation and Sungkyunkwan University. He is currently an Associate Professor with Yonsei University. His research interests include deep learning, intelligent design methodologies, and system reliability in emerging architectures. He was nominated for the Best Paper Award at the 2013 IEEE VLSI Test Symposium. He was a recipient of the Korea Science and Engineering Foundation (KOSEF) Scholarship, in 2005. He received the Best Paper Award at the 2008 IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems and at the 2016 IEEE International SoC Design Conference.